

# BMC Evolutionary Biology

Research article

## High correlation between the turnover of nucleotides under mutational pressure and the DNA composition

Maria Kowalczyk<sup>1</sup>, Pawel Mackiewicz<sup>1</sup>, Dorota Mackiewicz<sup>1</sup>, Aleksandra Nowicka<sup>1</sup>, Malgorzata Dudkiewicz<sup>1</sup>, Mirosław R Dudek<sup>2</sup> and Stanisław Cebrat\*<sup>1</sup>

Address: <sup>1</sup>Institute of Microbiology, Wrocław University, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland and <sup>2</sup>Institute of Physics, Zielona Góra University, Poland

E-mail: Maria Kowalczyk - [kowal@microb.uni.wroc.pl](mailto:kowal@microb.uni.wroc.pl); Pawel Mackiewicz - [pamac@microb.uni.wroc.pl](mailto:pamac@microb.uni.wroc.pl); Dorota Mackiewicz - [dorota.s@microb.uni.wroc.pl](mailto:dorota.s@microb.uni.wroc.pl); Aleksandra Nowicka - [nowic@microb.uni.wroc.pl](mailto:nowic@microb.uni.wroc.pl); Malgorzata Dudkiewicz - [malgosia@merry.microb.uni.wroc.pl](mailto:malgosia@merry.microb.uni.wroc.pl); Mirosław R Dudek - [mdudek@proton.if.wsp.zgora.pl](mailto:mdudek@proton.if.wsp.zgora.pl); Stanisław Cebrat\* - [cebrat@microb.uni.wroc.pl](mailto:cebrat@microb.uni.wroc.pl)

\*Corresponding author

Published: 17 December 2001

Received: 10 September 2001

*BMC Evolutionary Biology* 2001, 1:13

Accepted: 17 December 2001

This article is available from: <http://www.biomedcentral.com/1471-2148/1/13>

© 2001 Kowalczyk et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any non-commercial purpose, provided this notice is preserved along with the article's original URL. For commercial use, contact [info@biomedcentral.com](mailto:info@biomedcentral.com)

### Abstract

**Background:** Any DNA sequence is a result of compromise between the selection and mutation pressures exerted on it during evolution. It is difficult to estimate the relative influence of each of these pressures on the rate of accumulation of substitutions. However, it is important to discriminate between the effect of mutations, and the effect of selection, when studying the phylogenetic relations between taxa.

**Results:** We have tested in computer simulations, and analytically, the available substitution matrices for many genomes, and we have found that DNA strands in equilibrium under mutational pressure have unique feature: the fraction of each type of nucleotide is linearly dependent on the time needed for substitution of half of nucleotides of a given type, with a correlation coefficient close to 1. Substitution matrices found for sequences under selection pressure do not have this property. A substitution matrix for the leading strand of the *Borrelia burgdorferi* genome, having reached equilibrium in computer simulation, gives a DNA sequence with nucleotide composition and asymmetry corresponding precisely to the third positions in codons of protein coding genes located on the leading strand.

**Conclusions:** Parameters of mutational pressure allow us to count DNA composition in equilibrium with this mutational pressure. Comparing any real DNA sequence with the sequence in equilibrium it is possible to estimate the distance between these sequences, which could be used as a measure of the selection pressure. Furthermore, the parameters of the mutational pressure enable direct estimation of the relative mutation rates in any DNA sequence in the studied genome.

### Background

Nucleotide sequences of all contemporary genomes are results of compromise between mutational pressure and

selection [1]. Many mutations which took place in the past have been eliminated by genetic death. Even so, there are differences in the nucleotide composition of protein

coding sequences and intergenic sequences – it is very difficult to discriminate between the effects of selection and mutations on their composition. Furthermore, many prokaryotic genomes have very asymmetric nucleotide composition of chromosomes [e.g. [2–9]]. Strand composition of DNA depends on the role which the strand plays in the replication process – leading or lagging. Usually the leading strand is richer in Guanine (G) than in Cytosine (C) and richer in Thymine (T) than in Adenine (A). The replication-associated mutational pressure is thought to be the most probable cause of this asymmetry [10–12].

Analyses of long range correlations in DNA sequences revealed that in the intergenic sequences a very strong triplet signal can be detected [13,14]. This signal can be created by fragments of coding sequences transferred into intergenic space by recombination mechanisms. Since the nucleotide compositions of the first, the second and the third nucleotide positions in coding sequences are strongly correlated, these correlations are seen even in some noncoding intergenic sequences. We have assumed that some intergenic sequences have derived from coding sequences and could freely accumulate mutations with frequencies determined by the replication-associated mutational pressure. If the time of divergence has not been very long, the homology between the intergenic sequences and their original protein coding sequences can be found (these original coding sequences we have called the reference sequences).

We have made an assumption that mutations have been accumulated only in the intergenic sequences and not in the reference sequences, which is not exactly true, but which enabled accomplishing our studies. This assumption could give a good approximation of mutational pressure exerted on intergenic sequences. Many other authors, who have constructed matrices of substitutions using the mutations accumulated in pseudogene sequences have made the same assumptions [15,16]. Such an assumption could give higher estimated mutational rate than the real one. Nevertheless, the substitution rates in the matrices are described as a relative values thus, it should not change the values in the matrix.

We have chosen for our analyses the *B. burgdorferi* genome because there are many premises indicating that this genome is in the steady state. The *B. burgdorferi* genome is very asymmetric, which suggests its structural conservation [17]. There are not many inversions of genes between the leading and lagging strands or the mutational pressure has had enough time to make the inverted genes resemble the genes of the new strand [18]. The nucleotide composition of the third positions in codons testifies for the very conserved structure of chromosome. These positions follow precisely the sign of the asymmetry of intergenic se-

quences and, the third positions of Open Reading Frames (ORFs) situated on the leading and lagging strands have precisely mirror asymmetry, which is even stronger than that of intergenic sequences [17]. This paradox could be explained assuming that the highly degenerated third positions have accumulated more neutral or near neutral mutations introduced by the replication-associated processes because they stay at their positions longer than intergenic sequences. There are constraints for inversions of coding sequences but no constraints for inversions of intergenic sequences. Thus, some newly inverted intergenic sequences could complement the asymmetry of the "new host" strand.

## Results and Discussion

### Testing the table of substitutions and verifying the assumptions

Once having experimentally found the rates of all types of substitutions (Table 1, *Borrelia burgdorferi* Table of Substitutions (BbTS)), we were able to test these data and to verify our previous assumptions. In equilibrium, the number of a given nucleotide substituted by other nucleotides should be balanced by the number of that nucleotide substituting the other nucleotides. The following four equations should be fulfilled:

$$N_{A>G} + N_{A>C} + N_{A>T} = N_{G>A} + N_{C>A} + N_{T>A} \quad (1)$$

$$N_{G>A} + N_{G>C} + N_{G>T} = N_{A>G} + N_{C>G} + N_{T>G} \quad (2)$$

$$N_{C>A} + N_{C>G} + N_{C>T} = N_{A>C} + N_{G>C} + N_{T>C} \quad (3)$$

$$N_{T>A} + N_{T>G} + N_{T>C} = N_{A>T} + N_{G>T} + N_{C>T} \quad (4)$$

where  $N_{A>G} = N_A \cdot p(N_{A>G})$  and where  $p(N_{A>G})$  is the probability of substitutions of A by G, taken from the BbTS, (other symbols – respectively).

Note that there are numbers, not frequencies in the equations. Fulfilling these equations means that the nucleotide composition of the sequences submitted to the mutational pressure determined by the parameters of BbTS is in equilibrium. We have assumed that in the case of the *B. burgdorferi* genome the best approximation of such sequences is the composition of the third positions of codons of ORFs, as has been argued in the Introduction section. Thus, the nucleotide composition of this set of nucleotides should not change significantly under such mutational pressure. To prove that, we simulated the mutational pressure on the sequence of the same composition as described previously [19] and after 10,000 Monte Carlo Steps (MCS), when the sequence was in equilibrium, we compared it to the sequence before the simulation. The ratios of nucleotides were 0.994, 1.008, 0.992 and 0.988 for A, T, G, and C, respectively (note that the ra-

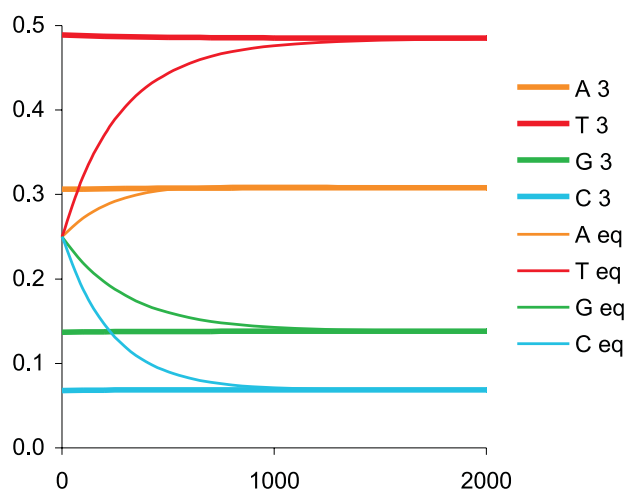
**Table 1: Tables of substitutions, DNA composition in the equilibrium with the mutational pressure and half times of nucleotide substitutions.**

Substitution	BbTS real	BbTS artificial	pseud 1	pseud 2	pseud m	pseud h	mtdros 4d	mtdros 3p	mtDNA 1, 2p
A>T	0,103	0,222	0,047	0,024	0,017	0,087	0,124	0,019	0,069
A>G	0,067	0,111	0,050	0,138	0,068	0,149	0,013	0,148	0,082
A>C	0,023	0,016	0,094	0,031	0,068	0,000	0,029	0,005	0,094
T>A	0,066	0,189	0,044	0,024	0,057	0,025	0,004	0,021	0,054
T>G	0,035	0,002	0,082	0,030	0,114	0,013	0,075	0,005	0,016
T>C	0,035	0,002	0,033	0,126	0,095	0,025	0,379	0,086	0,182
G>A	0,164	0,023	0,210	0,214	0,166	0,177	0,056	0,449	0,118
G>T	0,116	0,178	0,072	0,047	0,030	0,076	0,000	0,041	0,030
G>C	0,015	0,056	0,053	0,051	0,060	0,050	0,000	0,000	0,031
C>A	0,070	0,186	0,065	0,052	0,076	0,052	0,288	0,000	0,087
C>T	0,261	0,010	0,210	0,211	0,172	0,312	0,000	0,227	0,216
C>G	0,047	0,004	0,042	0,054	0,076	0,035	0,124	0,000	0,020
DNA composition									
A	30,8	30,8	31,3	30,8	41,0	17,0	27,4	29,8	24,7
T	48,5	48,5	37,2	30,8	17,1	65,6	50,1	43,6	31,5
G	13,8	13,8	15,8	19,9	23,9	11,7	8,3	9,4	17,2
C	6,9	6,9	15,8	18,5	18,1	5,6	14,3	17,2	26,6
Half time of substitution									
$\tau_A$	361	199	364	360	450	294	412	404	283
$\tau_T$	513	359	437	385	262	1103	637	622	274
$\tau_G$	236	269	207	223	270	229	159	141	388
$\tau_C$	183	346	219	219	214	174	240	306	214
Correlation coeff.	0,999	0,072	0,998	0,991	0,967	0,998	0,997	0,988	-0,764

Explanations: BbTS real – the table of substitutions in the *B. burgdorferi* genome on the leading strand estimated as described in the Method section; BbTS artificial – one of the possible tables which produce the DNA asymmetry and composition like BbTS real; pseud 1 – data for mammal pseudo-gene sequences (Li, Wu, and Luo 1984); pseud 2 – data for the psi-eta-globin pseudogenes of primates (Yang 1994); pseud m and pseud h – data for LDH-A pseudogenes of mouse and human respectively (Fukasawa et al. 1986); mtdros 4d and mtdros 3p – data for cytochrome b and NADH dehydrogenase subunit 1 genes of *Drosophila* mtDNA for four-fold degenerate sites and the third codon positions (Tamura 1992); mtDNA 1, 2p – data for the first and the second codon positions of vertebrate mitochondrial genes (Zhang 1999) Notes: The last column represents data for substitutions in mitochondrial sequences under strong selection pressure;  $p_{mut}$  – parameter for  $\tau$  counting equals 0,01. For more explanations see text.

tios do not sum to 1 because they are not weighted). There are no significant changes in nucleotide composition of the third positions after the prolonged exposition to the mutational pressure described by BbTS (Chi square test,  $p = 0.99987$ ). Thus, BbTS generates DNA sequence with nucleotide composition corresponding precisely to the nucleotide composition of the third codon positions. In Fig. 1 we have shown the evolution of two DNA sequences, of which one originally had equimolar nucleotide composition and the other one – the nucleotide composition of

the third codon positions of ORFs from the leading strand. Both sequences reach the same final nucleotide composition. Furthermore, a sequence obtained after long evolution in computer has very similar asymmetry in terms of GC skew and AT skew as the sequence of the third codon positions before evolution. GC skew is  $[G-C]/[G+C]$  and AT skew is  $[A-T]/[A+T]$ . The AT skew is -0.23 and -0.22 for the sequences before and after simulation, respectively. The GC skew is 0.34 for the sequences before and after simulation. Note that the most frequent substi-



**Figure 1**

Evolution of DNA sequences under the mutational pressure described by the "real BbTS". Light lines indicate the fractions of nucleotides in the sequence which initially has been composed of equal numbers of each nucleotide. Bold lines show the fractions of nucleotides in a sequence of nucleotide composition of the third positions in codons of the *B. burgdorferi* coding sequences from the leading strand. x-axis – the number of Monte Carlo Steps (MCS), y-axis – fraction of nucleotides in the evolving strand.

tution is C→T transition, which is in agreement with the cytosine deamination theory (see ref. 10 and references therein), and the average transition frequency is twofold higher than transversion frequency.

#### Properties of the substitution matrices

Let us consider only nucleotides existing in the original sequence, which is already in the steady state. It is trivial that the substitution of each of the four nucleotides will follow exactly the same rules as a decay of radioactive isotopes with characteristic for each nucleotide "half time of substitutions" ( $\tau_A$ ,  $\tau_G$ ,  $\tau_T$ ,  $\tau_C$  for A, G, T, and C, respectively) determined by the sum of probabilities of substitutions of a given nucleotide by the other three nucleotides. In a more formal language:

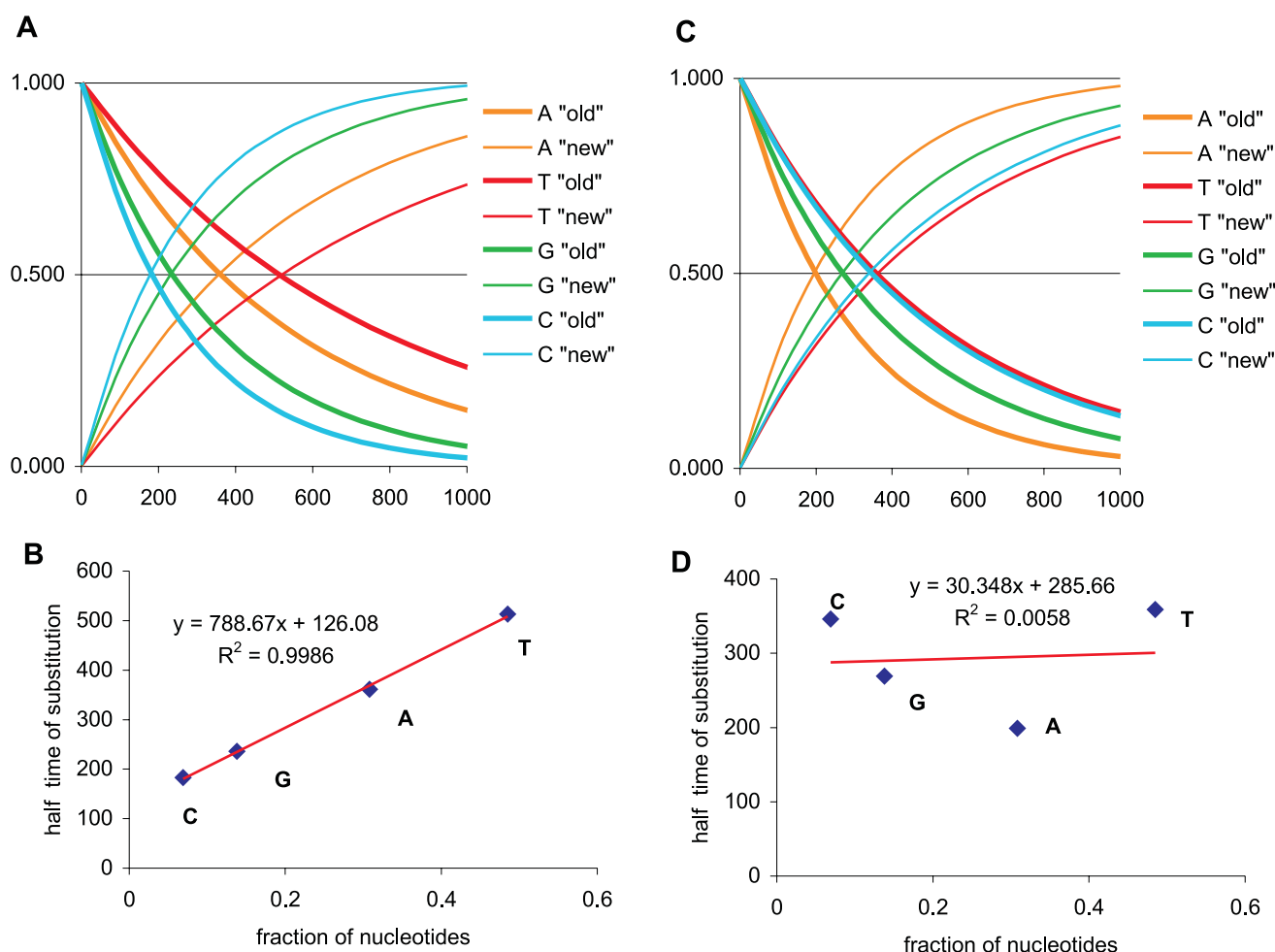
$\tau_A = \ln 2 / (p_{mut} * (p(A>G) + p(A>T) + p(A>C)))$ ; (symbols for nucleotides other than A-respectively), where  $p_{mut}$  is a parameter which denotes the overall rate of mutations and does not influence the ratios between  $\tau$  for different nucleotides.

It is also trivial that in the equilibrium, the fraction of a nucleotide which has been substituted is exactly the same as the fraction of this very nucleotide substituting the other ones (left sides of equations 1–4). Thus, after the half time of substitutions the ratio between the "old" nucle-

otides and "new" nucleotides is 1:1 (see Fig. 2A and Fig. 2C). This is a general property of any table of substitutions in the equilibrium state. But BbTS has another property: the half time of substitutions is precisely correlated with the frequency of the given nucleotide in the sequence in equilibrium with the correlation coefficient equaling 0.999 ( $p = 0.0007$ , Fig. 2B). This is not just a feature of any matrix of substitutions. We have tried to find analytically a table of substitutions which would generate a DNA sequence of the nucleotide composition of the analysed sequence of *B. burgdorferi* [see also [19]]. One of such tables is presented in Table 1. This "artificial table" generates a DNA sequence of the same nucleotide composition as BbTS does, but the correlation coefficient between the half time of substitutions and the fraction of nucleotides in the sequence is close to zero (Fig. 2C and 2D).

We claim that the mutational pressure leading to the nucleotide substitutions is extremely highly correlated with the DNA composition of the genome in such a way that the higher substitution turnover of a nucleotide determines the lower fraction of this nucleotide in the DNA sequence.

It seems very unbelievable that such correlation in the *B. burgdorferi* genome has happened accidentally. We have tested many other tables of substitutions which had been published for different genomes and different sequences (data collected in Table 1). As long as such matrices describe the substitutions for sequences which are not under the selection pressure (i.e. pseudogenes or the third positions in codons), they follow the same rule, with extremely high correlation between  $\tau_N$  and the fraction of the nucleotide *N* in the DNA sequence in equilibrium (all correlations were statistically significant). This rule is true for asymmetric DNA like in the *B. burgdorferi* genome and for much less biased eukaryotic DNA. Matrices found for the third positions in the four-fold degenerated codons in *Drosophila* mitochondrial DNA [20] fulfil this rule more precisely than for all third positions in codons in that organelle's genome (the same results were obtained for matrices of primates' mtDNA published by [21], data not shown). These differences could be expected if some mutations in the third positions, leading to amino acid substitutions are not neutral. Furthermore, in some instances, for example for the table describing substitution rates in sequences under strong selection [22], we have not found the correlation between  $\tau_N$  and the fraction of nucleotide *N* (see the last column in Table 1). That supports the hypothesis that the rule is a specific property of the pure mutational pressure. One can also notice that matrices found by analysis of substitutions into different pseudogenes in the same organism or in very closely related organisms give a different DNA composition in equilibrium, which

**Figure 2**

The rate of substitution of nucleotides in the DNA sequence in equilibrium under the mutational pressure described by: A – the "real BbTS" and, C – the "artificial BbTS". Bold lines show the fractions of nucleotides which have not been substituted yet, light lines indicate the fractions of nucleotides which appeared when substituting other nucleotides. Plots B and D represent the relations between half time of substitutions and the sizes of nucleotide fractions for sequences in equilibrium under the mutational pressure of "real BbTS" and "artificial BbTS" respectively. See also text and description for Fig. 1.

supports the thesis that the mutational pressure varies for different regions of the same eukaryotic genome [23–25].

We have no clear answer for the question: what selection forces have tuned the mutational pressure in such a way that it follows the strict rules for sequences released from selection. It is logical that nucleotides with higher turnover destabilise the genetic information and selection would tend to eliminate them from the DNA molecule. On the other hand, a lower frequency of a nucleotide gives it a higher informative value while at the same time the deviation from the equimolar fractions of nucleotides in DNA diminishes the coding capacity of the whole molecule. Perhaps mathematical analysis of this phenomena,

taking into considerations the properties of the universal genetic code, will show that the optimum for information transfer by the DNA molecule is just at such points. Further studies would show other properties of these strategic points where  $\tau$  determines very specific balance between the DNA composition and mutational pressure.

The implications of such evolutionary established relations between the DNA composition and the turnover rates of nucleotides would have a great impact on the understanding of the genomes evolution itself. It gives the possibility of estimating the relations between the mutational pressure exerted on specific nucleotides of each genome analytically i.e. by simple computing the nucleotide

composition of sequences which are not subject to selection pressure. Having the mutational pressure in terms of nucleotide turnover, one can estimate the selection pressure exerted on any sequence or position in codons. For example, see what would be the fate of the first positions of ORFs from the leading strand of the *B. burgdorferi* genome under the BbTS molecular pressure, without selection (Fig. 3). Note that the half time of substitutions of each nucleotide is the same as for other sequences under such mutational pressure, but the rate of appearing of new nucleotides is different and the composition of the sequence would change non-linearly during evolution. It is also simple to count, from the results of computer simulations, the corrections for multiple substitutions and reversions, which is important for estimating the real divergence time. It is clear that such corrections should be counted considering different contributions of each nucleotide turnover in the overall frequencies of multiple substitutions. Having precise mutational pressure one can predict not only the selection pressure but also find the history of the sequence.

Keeping in mind the precise relations between the fraction of nucleotide and its turnover time, the symmetric DNA (with A=T and G=C) is a specific case where the turnover times of nucleotides in pairs equal each other. Posing a question of which type of substitution should be blamed for the DNA asymmetry makes sense for the mutational pressure exerted on the DNA released from the selection pressure. Now a simple test for such a mutational pressure is available – it should generate the DNA in equilibrium whose nucleotide composition fulfils the rule of linear interdependence between the sizes of the nucleotide fractions and their turnover times.

## Conclusions

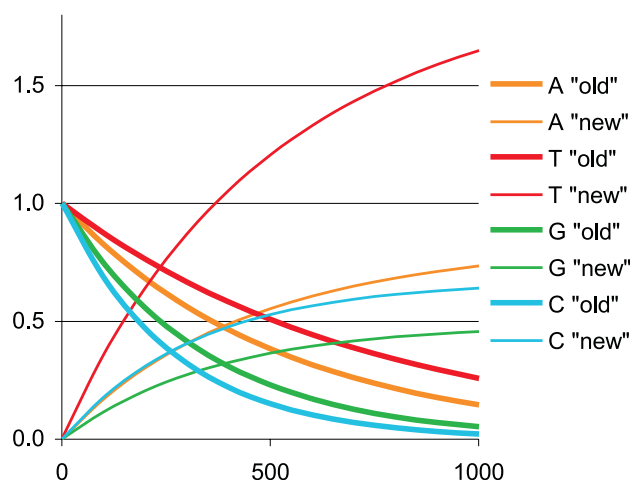
Substitution matrices enable counting the DNA composition in equilibrium with a given mutational pressure. It is possible to test if a given substitution matrix is the pure mutational matrix or if it is "contaminated" with the effects of selection. The difference between the DNA composition in equilibrium with mutational pressure and a DNA sequence under both mutational and selection pressures allows for estimation of the effect of selection pressure exerted on the particular sequence.

## Materials and Methods

### Construction of the substitution table

To estimate the frequency of substitutions, we have analysed the differences between coding sequences of the *B. burgdorferi* genome and sequences homologous to them found in the intergenic regions. For the data, see Additional file 1. The sequence of the *B. burgdorferi* genome [26] was downloaded from [http://www.ncbi.nlm.nih.gov]. To accomplish our analysis, we extracted all intergenic se-

quences longer than 90 nucleotides. We translated them into amino acid sequences in all six reading frames. The amber and ochre stop codons were translated for tyrosine residues and opal for tryptophan. Then we searched data bases for homology with the *B. burgdorferi* protein sequences using FASTA program [27]. For detailed amino acid alignment data see Additional file 2. After selecting homologues (with  $E < 0.05$ ) whose previously (presumably) coding strands were duplicated on the leading strand, we made alignments of nucleotide sequences of these intergenic sequences with the reference ORFs' sequences using CLUSTAL X programme [28] and we counted the nucleotide substitutions. The number of the analysed alignments sites was 3737 and the average number of substitutions per site – 0.46. For detailed nucleotide alignment data see Additional file 3. The observed numbers of nucleotide substitutions from nucleotide  $i$  to  $j$  (where  $i$  to  $j$  stand for A, T, G or C, and  $i \neq j$ ) were converted to relative substitution frequencies according to Gojobori, Li, and Gaur [29] and Francino and Ochman [30]. That allowed us to count the frequency of each of the twelve possible substitutions on the leading strand. Since the observed substitution rates were different for each of the four nucleotides, we introduced corrections for multiple substitutions and reversions for each type of the substituted nucleotide instead of one general correction. It means that we have counted for each type of nucleotide the fraction of substituted (observed number) and used it for estimat-



**Figure 3**

Changes in the nucleotide composition of the first positions in codons of *B. burgdorferi* coding sequences from the leading strand under BbTS mutational pressure. Descriptions as for Fig. 1 and 2. Note that the fractions of nucleotides which have not been substituted are exactly as in Fig. 2A but the fractions of nucleotides which substituted other nucleotides are far from being symmetrical to the first ones.

ing the corrected substitution number according to Kimura's formula [31]. The frequencies of substitutions, normalised in such a way that the sum of all 12 frequencies equals 1, are shown in Table 1 (Table of Substitutions – BbTS).

### Computer simulations

Computer simulations were performed on the DNA sequences corresponding to the real protein coding sequences of *B. burgdorferi* leading strand or DNA random sequences generated by computer. In the first case all ORFs longer than 100 codons situated on leading strand of *Borrelia* genome were spliced together. In the second case the DNA sequence were constructed by computer by drawing the consecutive nucleotides with a probability described by the assumed composition of this artificial sequence.

During the simulation of mutational pressure, in one Monte Carlo Step (MCS) each nucleotide in the sequence was drawn with the probability equalling  $p_{mut}$ , then it is substituted with the probability described by the substitution matrix. Note that the nucleotide drawn for substitution not necessarily has to be substituted. After each MCS each substitution was counted in the specific type of substitution and additionally the evolving sequence was compared with the original sequence and the accumulated substitution were counted. This allowed us to measure not only the divergence rate but also the mutational rate subdivided on 12 different classes. The dynamic of substitution was also counted analytically using the equations describing the first order reaction rate. For more details on computing methods see Kowalczyk et al. [19,32].

### Abbreviation

BbTS – table of probabilities of nucleotide substitutions in the *Borrelia burgdorferi* leading DNA strand.

### Additional material

#### Additional file 1

Contains information on intergenic sequences and homologous ORFs used to obtain table of substitutions. This file is to be opened in excel with tab as separator.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-1-13-S1.xls]

#### Additional file 2

Contains the relevant fragments of letters from FASTA with amino acid alignments.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-1-13-S2.txt]

#### Additional file 3

Contains nucleotide alignments by CLUSTALX program.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-1-13-S3.txt]

### Acknowledgements

Funding sources: This work was supported by The State Committee for Scientific Research, grant numbers 6 P04A 025-18 and 6 P04A 016 20. P.M. was supported by Foundation for Polish Science.

### References

1. Sueoka N: **On the genetic basis of variation and heterogeneity of DNA base composition.** *Proc Natl Acad Sci USA* 1962, **48**:582-591
2. Lobry JR: **Asymmetric substitution patterns in the two DNA strands of bacteria.** *Mol Biol Evol* 1996, **13**:660-665
3. Freeman JM, Plasterer TN, Smith TF, Mohr SC: **Patterns of genome organization in bacteria.** *Science* 1998, **279**:1827-1830
4. Grigoriev A: **Analyzing genomes with cumulative skew diagrams.** *Nucleic Acids Res* 1998, **26**:2286-2290
5. McLean MJ, Wolfe KH, Devine K: **Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes.** *J Mol Evol* 1998, **47**:691-696
6. Mrazek J, Karlin S: **Strand compositional asymmetry in bacterial and large viral genomes.** *Proc Natl Acad Sci USA* 1998, **95**:3720-3725
7. Francino MP, Ochman H: **A comparative genomics approach to DNA asymmetry.** *Ann N Y Acad Sci* 1999, **870**:428-431
8. Mackiewicz P, Gierlik A, Kowalczyk M, Dudek MR, Cebrat S: **Asymmetry of nucleotide composition of prokaryotic chromosomes.** *J Appl Genet* 1999, **40**:1-14
9. Rocha EP, Danchin A, Viari A: **Universal replication biases in bacteria.** *Mol Microbiol* 1999, **32**:11-16
10. Frank AC, Lobry JR: **Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms.** *Gene* 1999, **238**:65-77
11. Mackiewicz P, Gierlik A, Kowalczyk M, Dudek MR, Cebrat S: **How does replication-associated mutational pressure influence amino acid composition of proteins?** *Genome Res* 1999, **9**:409-416
12. Tillier ER, Collins RA: **The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes.** *J Mol Evol* 2000, **50**:249-257
13. Voss R: **Evolution of long-range fractal correlations and 1/f noise in DNA base sequences.** *Phys Rev Lett* 1992, **68**:3805-3808
14. Gierlik A, Mackiewicz P, Kowalczyk M, Dudek MR, Cebrat S: **Some hints on Open Reading Frame statistics – how ORF length depends on selection.** *Int J Modern Phys C* 1999, **10**:635-643
15. Li WH, Wu CI, Luo CC: **Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications.** *J Mol Evol* 1984, **21**:58-71
16. Yang Z: **Estimating the pattern of nucleotide substitution.** *J Mol Evol* 1994, **39**:105-111
17. Mackiewicz P, Gierlik A, Kowalczyk M, Szczepanik D, Dudek MR, Cebrat S: **Mechanisms generating long-range correlation in nucleotide composition of the *Borrelia burgdorferi*.** *Physica A* 1999, **273**:103-115
18. Mackiewicz P, Mackiewicz D, Gierlik A, Kowalczyk M, Nowicka A, Dudkiewicz M, Dudek MR, Cebrat S: **The Differential Killing of Genes by Inversions in Prokaryotic Genomes.** *J Mol Evol* 2001, **53**:615-621
19. Kowalczyk M, Gierlik A, Mackiewicz P, Cebrat S, Dudek MR: **Optimization of gene sequences under constant mutational pressure and selection.** *Physica A* 1999, **273**:116-131
20. Tamura K: **The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA.** *Mol Biol Evol* 1992, **9**:814-825



21. Adachi J, Hasegawa M: **Tempo and mode of synonymous substitutions in mitochondrial DNA of primates.** *Mol Biol Evol* 1996, **13**:200-208
22. Zhang J: **Performance of likelihood ratio tests of evolutionary hypotheses under inadequate substitution models.** *Mol Biol Evol* 1999, **16**:868-875
23. Filipinski J: **Why the rate of silent codon substitutions is variable within a vertebrate's genome.** *J Theor Biol* 1988, **134**:159-164
24. Wolfe KH, Sharp PM, Li WH: **Mutation rates differ among regions of the mammalian genome.** *Nature* 1989, **337**:283-285
25. Matassi G, Sharp PM, Gautier C: **Chromosomal location effects on gene sequence evolution in mammals.** *Curr Biol* 1999, **9**:786-791
26. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, et al: **Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.** *Nature* 1997, **390**:580-586
27. Pearson VVR, Lipman DJ: **Improved tools for biological sequence comparison.** *Proc Natl Acad Sci USA* 1988, **85**:2444-2448
28. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ: **Multiple sequence alignment with Clustal X.** *Trends Biochem Sci* 1988, **23**:403-405
29. Gojobori T, Li WH, Graur D: **Patterns of nucleotide substitution in pseudogenes and functional genes.** *J Mol Evol* 1982, **18**:360-369
30. Francino MP, Ochman H: **Strand symmetry around the beta-globin origin of replication in primates.** *Mol Biol Evol* 2000, **17**:416-422
31. Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.** *J Mol Evol* 1980, **16**:111-120
32. Kowalczyk M, Mackiewicz P, Mackiewicz D, Nowicka A, Dudkiewicz M, Dudek MR, Cebur S: **Multiple base substitution corrections in DNA sequence evolution.** *Int J Modern Phys C* 2001, **12**:1043-1053

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



**BioMedcentral.com**

Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

[editorial@biomedcentral.com](mailto:editorial@biomedcentral.com)